

WHAT IS CLAIMED IS:

- 1 1. A method of creating a library of DNA sequences, said method
2 comprising:
 - 3 a) providing a DNA sequence that encodes a protein of interest;
 - 4 b) providing a probability matrix for the protein;
 - 5 c) providing a constraint vector for the protein;
 - 6 d) applying the constraint vector to the probability matrix to produce a
7 substitution scheme recommending substitutions at at least two residues in the protein;
8 and
9 e) creating a library of DNA sequences incorporating changes in the
10 DNA sequence that produce the recommended substitutions.
- 1 2. The method of claim 1, wherein said protein is selected from the
2 group consisting of an esterase, dehydrogenase and hydrolase.
- 1 3. The method of claim 2, wherein said protein is selected from the
2 group consisting of a protease, cellulase, lipase, hemicellulase, laccase, and amylase.
- 1 4. The method of claim 1, wherein said protein is selected from the
2 group consisting of a transcription factor, growth factor, antibody, interleukin, antigen,
3 and receptor.
- 1 5. The method of claim 1, wherein the probability matrix is based on
2 structural characteristics selected from the group consisting of conservative residues,
3 sequence alignments, three dimensional structure, residue environment, solvent
4 accessibility, residue chemistry, propensity for a particular secondary structure, and
5 combinations thereof.
- 1 6. The method of claim 1, wherein the constraint vector is based on
2 structural characteristics known to affect protein function selected from the group
3 consisting of proximity to the site of functionality, distance of α or β carbons, contact
4 with residues of interest, and contact with residues that contact the residue of interest.

- 1 7. The library of claim 1, wherein said library is a phage library.
- 1 8. A method for screening a library for a protein with an increase in a
- 2 property of interest, comprising:
 - 3 a) providing a probability matrix for a protein of interest;
 - 4 b) providing a constraint vector for the protein;
 - 5 c) applying the constraint vector to the probability matrix to produce a
 - 6 substitution scheme recommending substitutions at at least two residues in the protein;
 - 7 and
 - 8 d) creating a library of DNA sequences incorporating changes in the
 - 9 DNA sequence that produce the recommended substitutions; and
 - 10 e) screening the library for a protein with an increase in the property
 - 11 of interest.

- 1 9. The method of claim 8, further comprising identifying a protein
- 2 having an increase in the property of interest.

- 1 10. A protein produced by the method of claim 9.
- 1 11. A system for creating libraries of nucleic acid sequences that
- 2 encode variants of a protein, said system comprising:
 - 3 a) an initial nucleic acid sequence that encodes a desired protein;
 - 4 b) a probability matrix; and
 - 5 c) a constraint vector.
- 1 12. A method for improving a desired parameter of a protein of
- 2 interest, comprising:
 - 3 a) providing a probability matrix for the desired protein;
 - 4 b) providing a constraint vector for the desired protein;
 - 5 c) applying the constraint vector to the probability matrix to produce a
 - 6 substitution scheme recommending substitutions at at least two residues in the protein;
 - 7 and
 - 8 d) creating a library of DNA sequences incorporating changes in the

9 DNA sequence that produce the recommended substitutions; and

10 e) measuring the parameter of interest for at least two members of
11 said library;

12 f) determining the sequence for at least two members of said library,
13 and

14 g) using sequence comparison and correlation analysis to determine
15 the contribution of mutations or combination of mutations on the parameter measured in
16 step e).

1 13. The method of claim 12, wherein the contribution of mutations
2 determined in step g) is used to generate a second library.

1 14. The method of claim 1, wherein a library comprising at least 25
2 unique DNA sequences is produced.

1 15. The method of claim 14, wherein a library comprising at least 100
2 unique DNA sequences is produced.

1 16. The method of claim 15, wherein a library comprising at least 250
2 unique DNA sequences is produced.

1 17. The method of claim 16, wherein a library comprising at least 1000
2 unique DNA sequences is produced.

1 18. The method of claim 17, wherein a library comprising at least 2500
2 unique DNA sequences is produced.

1 19. The method of claim 18, wherein a library comprising at least
2 10,000 unique DNA sequences is produced.

1 20. The method of claim 1, wherein a library of less than 10^9 unique
2 DNA sequences is produced.

1 21. The method of claim 20, wherein a library of less than 10^6 unique
2 DNA sequences is produced.

1 22. The method of claim 21, wherein a library of less than 10^5 unique
2 DNA sequences is produced.

1 23. The method of claim 1, wherein the probability matrix is an

2 algorithm.

1 24. The method of claim 1, wherein the probability matrix is generated
2 by a computer.

1 25. The method of claim 1, wherein the constraint vector is an
2 algorithm.

1 26. The method of claim 1, wherein the constraint vector is generated
2 by a computer.

1 27. The method of claim 1, wherein the constraint vector is applied to
2 the probability matrix using a computer.

1 28. The method of claim 1, wherein the probability matrix is
2 normalized.

1 29. The method of claim 1, wherein the DNA sequence is generated
2 from DNA shuffling.

1 30. The method of claim 9, further comprising using a DNA sequence
2 encoding the protein having an increase in the property of interest in a DNA shuffling
3 process.

1 31. A method of creating a library of DNA sequences, said method
2 comprising:

3 a) providing a substitution scheme produced by applying a constraint
4 vector to a probability matrix wherein the substitution scheme recommends substitutions
5 at at least two residues in a protein of interest; and

6 b) creating a library of DNA sequences incorporating substitutions in
7 a DNA sequence encoding the protein of interest to create a library comprising the
8 recommended substitutions.